PRODUCTION OF SYRINGYL LIGNIN IN GYMNOSPERMS

Field of the Invention

This application claims the benefit of U.S. Provisional Application number 60/033,381, filed December 16, 1996. The invention relates to the molecular modification of gymnosperms in order to cause the production of syringyl units during lignin biosynthesis and to production and propagation of gymnosperms containing syringyl lignin.

Background of the Invention

Lignin is a major part of the supportive structure of most woody plants including angiosperm and gymnosperm trees which in turn are the principal sources of fiber for making paper and cellulosic products. In order to liberate fibers from wood structure in a manner suitable for making many grades of paper, it is necessary to remove much of the lignin from the fiber/lignin network. Lignin is removed from wood chips by treatment of the chips in an alkaline solution at elevated temperatures and pressure in an initial step of papermaking processes. The rate of removal of lignin from wood of different tree species varies depending upon lignin structure. Three different lignin structures have been identified in trees: p-hydroxyphenyl, guaiacyl and syringyl, which are illustrated in Fig. 1.

Angiosperm species, such as *Liquidambar styraciflua* L. [sweetgum], have lignin composed of a mixture of guaiacyl and syringyl monomer units. In contrast, gymnosperm species such as *Pinus taeda* L. [loblolly pine] have lignin which is devoid of syringyl monomer units. Generally speaking, the rate of delignification in a pulping process is directly proportional to the amount of syringyl lignin present in the wood. The higher delignification rates associated with species having a greater proportion of syringyl lignin result in more efficient pulp mill operations since the mills make better use of energy and capital investment and the environmental impact is lessened due to a decrease in chemicals used for delignification.

It is therefore an object of the invention to provide gymnosperm species which are easier to delignify in pulping processes.

Another object of the invention is to provide gymnosperm species such as loblolly pine which contain syringyl lignin.

An additional object of the invention is to provide a method for modifying genes involved in ligning biosynthesis in gymnosperm species so that production of syringyl lignin is increased while production of guaiacyl lignin is suppressed.

Still another object of the invention is to produce whole gymnosperm plants containing genes which increase production of syringyl lignin and repress production of guaiacyl lignin.

Yet another object of the invention is to identify, isolate and/or clone those genes in angiosperms responsible for production of syringyl lignin.

A further object of the invention is to provide, in gymnosperms, genes which produce syringyl

lignin.

Another object of the invention is to provide a method for making an expression cassette insertable into a gymnosperm cell for the purpose of inducing formation of syringyl lignin in a gymnosperm plant derived from the cell.

Definitions

The term "promoter" refers to a DNA sequence in the 5' flanking region of a given gene which is involved in recognition and binding of RNA polymerase and other transcriptional proteins and is required to initiate DNA transcription in cells.

The term "constitutive promoter" refers to a promoter which activates transcription of a desired gene, and is commonly used in creation of an expression cassette designed for preliminary experiments relative to testing of gene function. An example of a constitutive promoter is 35S CaMV, available from Clonetech.

The term "expression cassette" refers to a double stranded DNA sequence which contains both promoters and genes such that expression of a given gene is acheived upon insertion of the expression cassette into a plant cell.

The term "plant" includes whole plants and portions of plants, including plant organs (e.g. roots, stems, leaves, etc.)

The term "angiosperm" refers to plants which produce seeds encased in an ovary. A specific example of an angiosperm is *Liquidambar styraciflua* (L.)[sweetgum]. The angiosperm sweetgum produces syringyl lignin.

The term "gymnosperm" refers to plants which produce naked seeds, that is, seeds which are not encased in an ovary. A specific example of a gymnosperm is *Pinus taeda* (L.)(loblolly pine]. The gymnosperm loblolly pine does not produce syringyl lignin.

Summary of the Invention

With regard to the above and other objects, the invention provides a method for inducing production of syringyl lignin in gymnosperms and to gymnosperms which contain syringyl lignin for improved delignification in the production of pulp for papermaking and other applications. In accordance with one of its aspects, the invention involves cloning an angiosperm DNA sequence which codes for enzymes involved in production of syringyl lignin monomer units, fusing the angiosperm DNA sequence to a lignin promoter region to form an expression cassette, and inserting the expression cassette into a gymnosperm genome.

Enzymes required for production of syringyl lignin in an angiosperm are obtained by deducing an amino acid sequence of the enzyme, extrapolating an mRNA sequence from the amino acid sequence, constructing a probe for the corresponding DNA sequence and cloning the DNA sequence which codes for the desired enzyme. A promoter region specific to a gymnosperm lignin biosynthesis

gene is identified by constructing a probe for a gymnosperm lignin biosynthesis gene, sequencing the 5' flanking region of the DNA which encodes the gymnosperm lignin biosynthesis gene to locate a promoter sequence, and then cloning that sequence.

An expression cassette is constructed by fusing the angiosperm syringyl lignin DNA sequence to the gymnosperm promoter DNA sequence. Alternatively, the angiosperm syringyl lignin DNA is fused to a constitutive promoter to form an expression cassette. The expression cassette is inserted into the gymnosperm genome to transform the gymnosperm genome. Cells containing the transformed genome are selected and used to produce a transformed gymnosperm plant containing syringyl lignin.

In accordance with the invention, the angiosperm gene sequences bi-OMT, 4CL, FA5H-1 and FA5H-2 have been determined and isolated as associated with production of syringyl lignin in sweetgum and lignin promoter regions for the gymnosperm loblolly pine have been determined to be the 5' flanking regions for the 4CL1B, 4CL3B and PAL gymnosperm lignin genes. Expression cassettes containing sequences of selected genes from sweetgum have been inserted into loblolly pine embryogenic cells and presence of sweetgum genes associated with production of syringyl lignin has been confirmed in daughter cells of the resulting loblolly pine embryogenic cells.

The invention therefore enables production of gymnosperms such as loblolly pine containing genes which code for production of syringyl lignin, to thereby produce in such species syringyl lignin in the wood structure for enhanced pulpability.

Brief Description of the Drawings

The above and other aspects of the invention will now be further described in the following detailed specification considered in conjunction with the following drawings in which:

Fig. 1 illustrates a generalized pathway for lignin synthesis; and

Fig. 2 illustrates a bifunctional-O-methyl transferase (bi-OMT) gene sequence involved in the production of syringyl lignin in an angiosperm (SEQ ID 3);

Fig. 3 illustrates a 4-coumarate CoA ligase (4CL) gene sequence involved in the production of syringyl lignin in an angiosperm (SEQ ID 4);

Fig. 4 illustrates a ferulic acid-5-hydroxylase (FA5H-1) gene sequence involved in the production of syringyl lignin in an angiosperm (SEQ ID 1);

Fig. 5 illustrates a ferulic acid-5-hydroxylase (FA5H-2) gene sequence involved in the production of syringyl lignin in an angiosperm (SEQ ID 2);

Fig. 6 illustrates nucleotide sequences of the 5' flanking region of the loblolly pine 4CL1B gene showing the location of regulatory elements for lignin biosynthesis (SEQ ID 6);

Fig. 7 illustrates nucleotide sequences of the 5' flanking region of the loblolly pine 4CL3B gene showing the location of regulatory elements for lignin biosynthesis (SEQ ID 7);

Fig. 8 illustrates nucleotide sequences of the 5' flanking region of loblolly pine PAL gene

showing the location of regulatory elements for lignin biosynthesis (SEQ ID 5);

Fig. 9 illustrates a PCR confirmation of the sweetgum FA5H-1 gene sequence in transgenic loblolly pine cells; and

Detailed Description of the Invention

In accordance with the invention, a method is provided for modifying a gymnosperm genome, such as the genome of a loblolly pine, so that syringyl lignin will be produced in the resulting plant, thereby enabling cellulosic fibers of the same to be more easily separated from lignin in a pulping process. In general, this is accomplished by fusing one or more angiosperm DNA sequences (referred to at times herein as the "ASL DNA sequences") which are involved in production of syringyl lignin to a gymnosperm lignin promoter region (referred to at times herein as the "GL promoter region") specific to genes involved in gymnosperm lignin biosynthesis to form a gymnosperm syringyl lignin expression cassette (referred to at times herein as the "GSL expression cassette"). Alternatively, the one or more ASL DNA sequences are fused to one or more constitutive promoters to form a GSL expression cassette.

The GSL expression cassette preferably also includes selectable marker genes which enable transformed cells to be differentiated from untransformed cells. The GSL expression cassette containing selectable marker genes is inserted into the gymnosperm genome and transformed cells are identified and selected, from which whole gymnosperm plants may be produced which exhibit production of syringyl lignin.

To suppress production of less preferred forms of lignin in gymnosperms, such as guaiacyl lignin, genes from the gymnosperm associated with production of these less preferred forms of lignin are identified, isolated and the DNA sequence coding for anti-sense mRNA (referred to at times herein as the "GL anti-sense sequence") for these genes is produced. The DNA sequence coding for anti-sense mRNA is then incorporated into the gymnosperm genome, which when expressed bind to the less preferred guaiacyl gymnosperm lignin mRNA, inactivating it.

Further features of these and various other steps and procedures associated with practice of the invention will now be described in more detail beginning with identification and isolation of ASL DNA sequences of interest for use in inducing production of syringyl lignin in a gymnosperm.

I. Determination of DNA Sequence For Genes Associated with Production of Syringyl Lignin

The general biosynthetic pathway for production of lignin has been postulated as shown in Fig. 1. From Fig. 1, it can be seen that the genes CCL, OMT and F5H (which is from the class of P450 genes) may play key roles in production of syringyl lignin in some plant species, but their specific contributions and mechanisms remain to be positively established. It is suspected that the CCL, OMT and F5H genes may have specific equivalents in a specific angiosperm, such as sweetgum. Accordingly, one aim of the present invention is to identify, sequence and clone specific genes of

interest from an angiosperm such as sweetgum which are involved in production of syringyl lignin and to then introduce those genes into the genome of a gymnosperm, such as loblolly pine, to induce production of syringyl lignin.

Genes of interest may be identified in various ways, depending on how much information about the gene is already known. Genes believed to be associated with production of syringyl lignin have already been sequenced from a few angiosperm species, viz, CCL and OMT.

DNA sequences of the various CCL and OMT genes are compared to each other to determine if there are conserved regions. Once the conserved regions of the DNA sequences are identified, oligo-dT primers homologous to the conserved sequences are synthesized. Reverse transcription of the DNA-free total RNA which was purified from sweetgum xylem tissue, followed by double PCR using gene-specific primers, enables production of probes for the CCL and OMT genes.

A sweetgum cDNA library is constructed in a host, such as lambda ZAPII, available from Stratagene, of LaJolla, CA, using poly(A) +RNA isolated from sweetgum xylem, according to the methods described by Bugos et al. (1995 Biotechniques 19:734-737). The above mentioned probes are used to assay the sweetgum cDNA library to locate cDNA which codes for enzymes involved in production of syringyl lignin. Once a syringyl lignin sequence is located, it is then cloned and sequenced according to known methods which are familiar to those of ordinary skill.

In accordance with the invention, two sweetgum syringyl lignin genes have been determined using the above-described technique. These genes have been designated 4CL and bi-OMT. The sequence obtained for the sweetgum syringyl lignin gene, designated bi-OMT, is illustrated in Fig. 2 (SEQ ID 3). The sequence obtained for the sweetgum syringyl lignin gene, designated 4CL, is illustrated in Fig. 3 (SEQ ID 4).

An alternative procedure was employed to identify the F5H equivalent genes in sweetgum. Because the DNA sequences for similar P450 genes from other plant species were known, probes for the P450 genes were designed based on the conserved regions found by comparing the known sequences for similar P450 genes. The known P450 sequences used for comparison include all plant P450 genes in the GenBank database. Primers were designed based on two highly conserved regions which are common to all known plant P450 genes. The primers were then used in a PCR reaction with the sweetgum cDNA library as a template. Once P450-like fragments were located, they were amplified using standard PCR techniques, cloned into a pBluescript vector available from Clonetech of Palo Alto, CA and transformed into a DH5α *E. coli* strain available from Gibco BRL of Gaithersburg, MD.

After *E. coli* colonies were tested in order to determine that they contained the P450-like DNA fragments, the fragments were sequenced. Several P450-like sequences were located in sweetgum using the above described technique. One P450-like sequence was sufficiently different from other

known P450 sequences to indicate that it represented a new P450 gene family. This potentially new P450 cDNA fragment was used as a probe to screen a full length clone from the sweetgum xylem library. This putative hydroxylase clone was designated FA5H-1. The sequence obtained for FA5H-1 is illustrated in Fig. 4 (SEQ ID 1).

II. Identification of GL Gene Promoter Regions

In order to locate gymnosperm lignin promoter regions, probes are developed to locate lignin genes. After the-gymnosperm lignin gene is located, the portion of DNA upstream from the gene is sequenced, preferably using the GenomeWalker Kit, available from Clonetech. The portion of DNA upstream from the lignin gene will generally contain the gymnosperm lignin promoter region.

Gymnosperm genes of interest include CCL-like genes and PAL-like genes, which are beleived to be involved in the production of lignin in gymnosperms. Preferred probe sequences are developed based on previously sequenced genes, which are available from the gene bank. The preferred gene bank accession numbers for the CCL-like genes include U39404 and U39405. A preferred gene bank accession number for a PAL-like gene is U39792. Probes for such genes are constructed according to methods familiar to those of ordinary skill in the art. A genomic DNA library is constructed and DNA fragments which code for gymnosperm lignin genes are then identified using the above mentioned probes. A preferred DNA library is obtained from the gymnosperm, *Pinus taeda* (L.)[Loblolly Pine], and a preferred host of the genomic library is Lambda DashII, available from Stratagene of LaJolla, CA.

Once the DNA fragments which code for the gymnosperm lignin genes are located, the genomic region upstream from the gymnosperm lignin gene (the 5'flanking region) was identified. This region contains the GL promoter. Three promoter regions were located from gymnosperm lignin biosynthesis genes. The first is the 5'flanking region of the loblolly pine 4CL1B gene, shown in Fig. 6 (SEQ ID 6). The second is the 5' flanking region of the loblolly pine gene 4CL3B, shown in Fig. 7 (SEQ ID 7). The third is the 5' flanking region of the loblolly pine gene PAL, shown in Fig. 8 (SEQ ID 5).

III. Fusing the GL Promoter Region to the ASL DNA Sequence

The next step of the process is to fuse the GL promoter region to the ASL DNA sequence to make a GSL expression cassette for insertion into the genome of a gymnosperm. This may be accomplished by standard techniques. In a preferred method, the GL promoter region is first cloned into a suitable vector. Preferred vectors are pGEM7Z, available from Promega, Madison, WI and SK available from Stratagene, of LaJolla, CA. After the promoter sequence is cloned into the vector, it is then released with suitable restriction enzymes. The ASL DNA sequence is released with the same restriction enzyme(s) and purified.

The GL promoter region sequence and the ASL DNA sequence are then ligated such as with T4

DNA ligase, available from Promega, to form the GSL expression cassette. Fusion of the GL and ASL DNA sequence is confirmed by restriction enzyme digestion and DNA sequencing. After confirmation of GL promoter-ASL DNA fusion, the GSL expression cassette is released from the original vector with suitable restriction enzymes and used in construction of vectors for plant transformation.

IV. Fusing the ASL DNA Sequence to a Constitutive Promoter Region

In an alternative embodiment, a standard constitutive promoter may be fused with the ASL DNA sequence to make a GSL expression cassette. For example, a standard constitutive promoter may be fused with FA5H-1 to form an expression cassette for insertion of FA5H-1 sequences into a gymnosperm genome. In addition, a standard constitutive promoter may be fused with FA5H-2 to form an expression cassette for insertion of FA5H-2 into a gymnosperm genome. A constitutive promoter for use in the invention is the double 35S promoter, available from Clonetech.

In the preferred practice of the invention using constitutive promters, a suitable vector such as pBi221, is digested XbaI and HindIII to release the 35S promoter. At the same time the vector pHygro, available from International Paper, was disgested by XbaI and HindIII to release the double 35S promoter. The double 35S promoter was ligated to the previously digested pBi221 vector to produce a new pBi221 with the double 35S promoter. This new pBi221 was digested with SacI and SmaI, to release the GUS fragment. The vector is next treated with T4 DNA polymerase to produce blunt ends and the vector is self-ligated. This vector is then further digested with BamHI and XbaI, available from Promega. After the pBi221 vector containing the constitutive promoter region has been prepared, lignin gene sequences are prepared for insertion into the pBi221 vector.

The coding regions of sweetgum FA5H-1 or FA5H-2 are amplified by PCR using primer with restriction sites incorporated in the 5' and 3' ends. In one example, an XbaI site was incorporated at the 5' end and a BamHI site was incorporated at the 3' end of the sweetgum FA5H-1 or FA5H-2 genes. After PCR, the FA5H-1 and FA5H-2 genes were separately cloned into a TA vector available from Invitrogen. The TA vectors containing the FA5H-1 and FA5H-2 genes, respectively, were digested by XbaI and BamHI to release the FA5H-1 or FA5H-2 sequences.

The p35SS vector, described above, and the isolated sweetgum FA5H-1 or FA5H-2 fragments were then ligated to make GLS expression cassettes containing the constiutive promoter.

V. Inserting the Expression Cassette into the Gymnosperm Genome

There are a number of methods by which the GSL expression cassette may be inserted into a target gymnosperm cell. One method of inserting the expression cassette into the gymnosperm is by micro-projectile bombardment of gymnosperm cells. For example, embryogenic tissue cultures of loblolly pine may be initiated from immature zygotic embryos. Tissue is maintained in an undifferentiated state on semi-solid proliferation medium. For transformation, embryogenic tissue is suspended in liquid proliferation medium. Cells are then sieved through, a preferably 40 mesh screen,

to separate small, densely cytoplasmic cells from large vacuolar cells.

After separation, a portion of the liquid cell suspension fraction is vacuum deposited onto filter paper and placed on semi-solid proliferation medium. The prepared gymnosperm target cells are then grown for several days on filter paper discs in a petri dish.

A 1:1 mixture of plasmid DNA containing the selectable marker expression cassette and plasmid DNA containing the FA5H-1 expression cassette may be precipitated with gold to form microprojectiles. The microprojectiles are rinsed in absolute ethanol and aliqots are dried onto a suitable macrocarrier such as the macrocarrier available from BioRad in Hercules, CA.

Prior to bombardment, embryogenic tissue is preferably desiccated under a sterile laminar-flow hood. The desiccated tissue is transferred to semi-solid proliferation medium. The prepared microprojectiles are accelerated from the macrocarrier into the desiccated target cells using a suitable apparatus such as a BioRad PDS-1000/HE particle gun. In a preferred method, each plate is bombarded once, rotated 180 degrees, and bombarded a second time. Preferred bombardment parameters are 1350 psi rupture disc pressure, 6 mm distance from the rupture disc to macrocarrier (gap distance), 1 cm macrocarrier travel distance, and 10 cm distance from macrocarrier stopping screen to culture plate (microcarrier travel distance). Tissue is then transferred to semi-solid proliferation medium containing a selection agent, such as hygromycin B, for two days after bombardment.

Other methods of inserting the GSL expression cassette include use of silicon carbide whiskers, transformed protoplasts, *Agrobacterium* vectors and electroporation.

VI. Identifying Transformed Cells

In general, insertion of the GSL expression cassette will typically be carried out in a mass of cells and it will be necessary to determine which cells harbor the recombinant DNA molecule containing the GSL expression cassette. Transformed cells are first identified by their ability to grow vigorously on a medium containing an antibiotic which is toxic to non-transformed cells. Preferred antibiotics are kanamycin and hygromycin B. Cells which grow vigorously on antibiotic containing medium are further tested for presence of either portions of the plasmid vector, the syringyl lignin genes in the GSL expression cassette; e.g. the angiosperm bi-OMT, 4CL, FA5H-1 or FA5H-2 gene, or by testing for presence of other fragments in the GSL expression cassette. Specific methods which can be used to test for presence of portions of the GSL expression cassette include Southern blotting with a labeled complementary probe or PCR amplification with specific complementary primers. In yet another approach, an expressed syringyl lignin enzyme can be detected by Western blotting with a specific antibody, or by assaying for a functional property such as the appearance of functional enzymatic activity.

VII. Production of a Gymnosperm Plant from the Transformed Gymnosperm Cell

Once transformed embryogenic cells of the gymnosperm have been identified, isolated and multiplied, they may be grown into plants. It is expected that all plants resulting from transformed cells will contain the GSL expression cassette in all their cells, and that wood in the secondary growth stage of the mature plant will be characterized by the presence of syringyl lignin.

Transgenic embryogenic cells are allowed to replicate and develop into a somatic embryo, which are then converted into a somatic seedling.

VIII. Identification, Production and Insertion of a GL mRNA Anti-Sense Sequence

In addition to adding ASL DNA sequences, anti-sense sequences may be incorporated into a gymnosperm genome, via GSL expression cassettes, in order to suppress formation of the less preferred native gymnosperm lignin. To this end, the gymnosperm lignin gene is first located and sequenced in order to determine its nucleotide sequence. Methods for locating and sequencing amino acids which have been previously discussed may be employed. For example, if the gymnosperm lignin gene has already been purified, standard sequencing methods may be employed to determine the DNA nucleic acid sequence.

If the gymnosperm lignin gene has not been purified and functionally similar DNA or mRNA sequences from similar species are known, those sequences may be compared to identify highly conserved regions and this information used as a basis for the construction of a probe. A gymnosperm cDNA or genomic library can be probed with the above mentioned sequences to locate the gymnosperm lignin cDNA or genomic DNA. Once the gymnosperm lignin DNA is located, it may be sequenced using standard sequencing methods.

After the DNA sequence has been obtained for a gymnosperm lignin sequence, the complementary anti-sense strand is constructed and incorporated into an expression cassette. For example, the GL mRNA anti-sense sequence may be fused to a promoter region to form an expression cassette as described above. In a preferred method, the GL mRNA anti-sense sequence is incorporated into the previously discussed GSL expression cassette which is inserted into the gymnosperm genome as described above.

IX. Inclusion of Cytochrome P450 Reductase (CPR) to Enhance Biosynthesis of Syringyl Lignin in Gymnosperms

In the absence of external cofactors such as NADPH (an electron donor in reductive biosyntheses), certain angiosperm lignin genes such as the FA5H genes may remain inactive or not acheive full or desired activity after insertion into the genome of a gymnosperm. Inactivity or insufficient activity can be determined by testing the resulting plant which contains the FA5H genes for the presence of syringyl lignin in secondary growth. It is known that cytochrome P450 reductase (CPR) may be involved in promoting certain reductive biochemical reactions, and may activate the desired

expression of genes in many plants. Accordingly, if it is desired to enhance the expression of the angiosperm syringyl lignin genes in the gymnosperm, CPR may be inserted in the gymnosperm genome. In order to express CPR, the DNA sequence of the enzyme is ligated to a constitutive promoter or, for a specific species such as loblolly pine, xylem-specific lignin promoters such as PAL, 4CL1B or 4CL3B to form an expression cassette. The expression cassette may then be inserted into the gymnosperm genome by various methods as described above.

X. Examples

The following non-limiting examples illustrate further aspects of the invention. In these examples, the angiosperm is *Liquidambar styraciflua* (L.)[sweetgum] and the gymnosperm is *Pinus taeda* (L.)[loblolly pine]. The nomenclature for the genes refered to in the examples is as follows:

Genes	Biochemical Name
4CL (angiosperm)	4-coumarate CoA ligase
bi-OMT (angiosperm)	bifunctional-O-methyl transferase
FA5H-1 (angiosperm)	ferulic acid-5-hydroxylase
FA5H-2 (angiosperm)	ferulic acid-5-hydroxylase
PAL (gymnosperm)	phenylalanine ammonia-lyase
4CL1B (gymnosperm)	4-coumarate CoA ligase
4CL3B (gymnosperm)	4-coumarate CoA ligase

Example 1 - Isolating and Sequencing bi-OMT and 4CL Genes from an Angiosperm

A cDNA library for Sweetgum was constructed in Lambda ZAPII, available from Stratagene, of LaJolla, CA, using poly(A) +RNA isolated from Sweetgum xylem tissue. Probes for bi-OMT and 4CL were obtained through reverse transcription of their mRNAs and followed by double PCR using gene-specific primers which were designed based on the OMT and CCL cDNA sequences obtained from similar genes cloned from other species.

Three primers were used for amplifying OMT fragments. One was an oligo-dT primer, bi-OMT, (which was cloned through modified differential display technique, as described below in Example 2) and the other two were degenerate primers, which were based on the conserved sequences of all known OMTs. The two degenerate primers were derived based on the following amino acid sequences:

5'- Gly Gly Met Ala Thr Tyr Cys Cys Ala Thr Thr Tyr Ala Ala Cys Ala Ala Gly Gly Cys-3' (primer #22) and

3'-Ala Ala Ala Gly Ala Gly Ala Gly Asn Ala Cys Asn Asn Ala Asn Asn Ala Asn Gly Ala-5'

(primer #23).

A 900 bp PCR product was produced when oligo-dT primer and primer #22 were used, and a 550 bp fragment was produced when primer numbers 22 and were used.

Three primers were used for amplifying CCL fragments. They were derived from the following amino acid sequences:

5'-Thr Thr Gly Gly Ala Thr Cys Cys Gly Gly Ile Ala Cys Ile Ala Cys Ile Gly Gly Ile Tyr Thr Ile Cys Cys Ile Ala Ala Arg Gly Gly-3' (primer R1S)

5'-Thr Thr Gly Gly Ala Thr Cys Cys Gly Thr Ile Gly Thr Ile Gly Cys Ile Cys Ala Arg Cys Ala Arg Gly Thr Ile Gly Ala Tyr Gly Gly-3' (primer H1S) and

3'-Cys Cys Ile Cys Thr Tyr Thr Ala Asp Ala Cys Arg Thr Ala Asp Gly Cys Ile Cys Cys Ala Gly Cys Thr Gly Thr Ala-5' (primer R2A)

R1S and H1S were both sense primers. Primer R2A was an anti-sense primer. A 650 bp fragment was produced if R1S and R2A primers were used and a 550 bp fragment was produced when primers H1S and R2A were used. The sequence of these three primers were derived from conserved sequences for plant CCLs.

The reverse transcription-double PCR cloning technique used for these examples consisted of adding 10 μm of DNA-free total RNA in 25μg DEPC-treated water to a microfuge tube. Next, the following solutions were added:

a. 5x Reverse transcript buffer 8.0μl,
b. 0.1 μM DDT 4.0 μl
c. 10 mM dNTP 2.0 μl
d. 100 μM oligo-dT primers 8.0 μl
e. Rnasin 2.0 μl
f. Superscript II 1.0 μl

After mixing, the tube was incubated at a temperature of 42° C for one (1) hour, followed by incubation at 70° C for fifteen (15) minutes. Forty (40) μl of 1N NaOH was added and the tube was further incubated at 68° C for twenty (20) minutes. After the incubation periods, 80 μl of 1N HCl was added to the reaction mixture. At the same time, 17 μl NaOAc, 5 μl glycogen and 768 μl of 100% ethanol were added and the reaction mixture was maintained at -80° C for 15 minutes in order to precipitate the cDNA. The precipitated cDNA was centrifuged at high speed at 4° C for 15 minutes. The resulting pellet was washed with 70% ethanol and then dried at room temperature, and then was dissolved in 20 μl of water.

The foregoing procedure produced purified cDNA which was used as a template to carry out first round PCR using primers #22 and oligo-dT for cloning OMT cDNA and primer R1S and R2A for

cloning 4CL cDNA. For the first round PCR, a master mix of 50μl for each reaction was prepared. Each 50μl mixture contained:

a. 10x buffer 5μl
b. 25 mM MgCl 5μl
c. 100 μM sense primer 1μl (primer #22 for OMT and primer R1S for CCL).
d. 100 μl anti-sense primer 1 μl (oligo-dT primer for OMT and R2A for CCL).
e. 10 mM dNTP 1 μl
f. Taq. DNA polymerase 0.5 μl

Of this master mix, 48 μl was added into a PCR tube containing 2 μl of cDNA for PCR. The tube was heated to 95° C for 45 seconds, 52°C for one minute and 72° C for two minutes. This temperature cycle was repeated for 40 cycles and the mixture was then held at 72° C for 10 minutes.

The cDNA fragments obtained from the first round of PCR were used as templates to perform the second round of PCR using primers 22 and 23 for cloning bi-OMT cDNA and primer H1S and R2A for cloning 4CL cDNA. The second round of PCR conditions were the same as the first round.

The desired cDNA fragment was then sub-cloned and sequenced. After the second round of PCR, the product with the predicted size was excised from the gel and ligated into a pUC19 vector, available from Clonetech, of Palo Alto, CA, and then transformed into DH5α, an *E. coli* strain , available from Gibco BRL, of Gaithersburg, MD. After the inserts had been checked for correct size, the colonies were isolated and plasmids were sequenced using a Sequenase kit available from USB, of Cleveland, OH. The sequences are shown in Fig. 2 (SEQ ID 3) and Fig. 3 (SEQ ID 4).

Example 2 - Alternative Isolation Method of Angiosperm bi-OMT gene

As previously mentioned, one bi-OMT clone was produced via modified differential display technique. This method is another type of reverse transcription-PCR, in which DNA-free total RNA was reverse transcribed using oligo-dT primers with a single base pair anchor to form cDNA. The oligo-dT primers used for reverse transcription of mRNA to synthesize cDNA were:

T11A: TTTTTTTTTTTTTTA,

T11C: TTTTTTTTTTTTTTC, and

T11G: TTTTTTTTTTTTTTG,

These cDNAs were then used as templates for radioactive PCR which was conducted in the presence of the same oligo-dT primers as listed above, a bi-OMT gene-specific primer and 35S-dATP. The OMT gene-specific primer was derived from the following amino acid sequence: 5'-Cys Cys Asn Gly Gly Asn Gly Gly Ser Ala Arg Gly Ala-3'.

The following PCR reaction solutions were combined in a microfuge tube:

a. H2O 9.2μl,
b. Taq Buffer 2.0μl
c. dNTP (25μM) 1.6μl

d. Primers (5 μM) 2 μl, for each primer
e. 35S-dATP 1μl
f. Taq. pol. 0.2μl
g. cDNA 2.0μl.

The tube was heated to a temperature of 94° C and held for 45 seconds, then at 37° C for 2' minutes and then 72°C for 45 seconds for forty cycles, followed by a final reaction at 72°C for 5 minutes.

The amplified products were fractionated on a denaturing polyacrylamide sequencing gel and autoradiography was used to identify and excise the fragments with a predicted size. The designed OMT gene-specific primer had a sequence conserved in a region toward the 3'-end of the OMT cDNA sequence. This primer, together with oligo-dT, was amplified into a OMT cDNA fragment of about 300 bp.

Three oligo-dTs with a single base pair of A, C or G, respectively, were used to pair with the OMT gene-specific primer. Eight potential OMT cDNA fragments with predicted sizes of about 300 bp were excised from the gels after several independent PCR rounds using different combinations of oligo-dT and OMT gene-specific oligo-nucleotides as primers.

The OMT cDNA fragments were then re-amplified. A Southern blot analysis was performed for the resulting cDNAs using a 360 base-pair, 32P radio-isotope labeled, aspen OMT cDNA 3'-end fragment as a probe to identify the cDNA fragments having a strong hybridization signal, under low stringency conditions. Eight fragments were identified. Out of these eight cDNA fragments, three were selected based on their high hybridization signal for sub-cloning and sequencing. One clone, LsOMT3'-1, (where the "Ls" prefix indicates that the clone was derived from the Liquidambar styraciflua (L.) genome) was confirmed to encode bi-OMT based on its high homology to other lignin-specific plant OMTs at both nucleotide and amino acid sequence levels.

A cDNA library was constructed in Lambda ZAP II, available from Stratagene, of LaJolla, CA, using 5mg poly(A)+RNA isolated from sweetgum xylem tissue. The primary library consisting of approximately 0.7 x 106 independent recombinants was amplified and approximately 105 plaque-forming-units (pfu) were screened using a homologous 550 base-pair probe. The hybridized filter was washed at high stringency (0.25 x SSC, 0.1% SDS, 65° C) conditions. The colony containing the bi-OMT fragment identified by the probe was eluted and the bi-OMT fragment was produced. The sequence as illustrated in Fig. 2 (SEQ ID 3) was obtained.

Example 3 - Isolating and Producing the DNA which codes for the Angiosperm FA5H-1 Gene

In order to find putative FA5H cDNA fragments as probes for cDNA library screening, a highly degenerated sense primer based on the amino acid sequence of 5'-Glu, Glu, Phe, Arg, Pro, Glu, Arg-3' was designed based on the conserved regions found in some plant P450 proteins. This conserved domain was located upstream of another highly conserved region in P450 proteins, which

had an amino acid sequence of 5'-Phe Gly Xaa Gly Xaa Xaa Cys Xaa Gly-3'. This primer was synthesized with the incorporation of an XboI restriction site to give a 26-base-pair oligomer with a nucleotide sequence of 5'ATG TGC AGT TTT TTT TTT TTT TTT TT-3'.

This primer and the oligo-dT-XhoI primer were then used to perform PCR reactions with the sweetgum cDNA library as a template. The cDNA library was constructed in Lambda ZAPII, available from Stratagene, of LaJolla, CA, using poly(a) +RNA isolated from Sweetgum xylem tissue. Amplified fragments of 300 to 600 bp were obtained. Because the designed primer was located upstream of the highly conserved P450 domain, this design distinguished whether the PCR products were P450 gene fragments depending on whether they contained the highly conserved amino acid domain.

All the fragments obtained from the PCR reaction were then cloned into a pUC19 vector, available from Stratagene, of LaJolla, CA, and transformed into a DH5α *E. coli* strain, available from Gibco BRL, of Gaithersburg, MD.

Twenty-four positive colonies were obtained and sequenced. Sequence analysis indicated four groupings withing the twenty-four colonies. One was C4H, one was an unknown P450 gene, and two did not belong to P450 genes. Homologies of P450 genes in different species are usually more than 80%. Because the homologies between the P450 gene families found here were around 40%, the sequence analysis indicated that a new P450 gene family was sequenced. Moreover, since this P450 cDNA was isolated from xylem tissue, it was highly probable that this P450 gene was FA5H-1.

The novel sweetgum P450 cDNA fragment was used as a probe to screen a full length cDNA encoding for FA5H-1. Once the FA5H-1 gene was located it was sequenced. The length of the FA5H-1 cDNA is 1707 bp and it contains 45 bp of 5' non-coding region and 135 bp of 3' non-coding region. The deduced amino acid sequence also indicates that this P450 cDNA has a hydrophobic core at the N-terminal, which could be regarded as a leader sequence for c-translational targeting to membranes during protein synthesis. At the C-terminal region, there is a heme binding domain that is characteristic of all P450 genes. The FA5H-1 sequence, as illustrated in Fig. 4 (SEQ ID 1), was produced, according to the above described methods.

Example 4 - Isolating and Producing the DNA which codes for the Angiosperm FA5H-2 Gene

By using similar strategy of synthesizing PCR primers from the published literature for hydroxylase genes in plants, another full length FA5H cDNA has been isolated that shows significant similarity with a putitive F5H clone from Arabidopsis (Meyers et al. 1996: PNAS 93, 6869-6874). This cloned cDNA, designated FA5H-2, contains 1883 bp and encodes an open reading frame of 511 amino acids. The amino acid similarity shared between Arabidopsis F5H and the FA5H-2 sweetgum clone is about 75%, indicating that the isolated clone belongs to the same class of cDNAs that encode a F5H protein, which has been shown to be functional by genetic complimentation in Arabidopsis.

To confirm the function of the FA5H-2 gene, it was expressed in *E.coli*, strain, DH5 alpha, via pQE vector preparation, according to directions available with the kit. A CO-Fe2+ binding assay was also performed to confirm the expression of FA5H-2 as a functional P450 gene. (Omura & Sato 1964, J. of Biochemistry 239: 2370-2378, Babriac et.al. 1991 Archives of Biochemistry and Biophysics 288:302-309). The CO-Fe2+ binding assay showed a peak at 450nm which indicates that FA5H-2 has been overexpressed as a functional P450 gene.

The FA5H-2 protein was further purified for production of antibodies in rabbits, and antibodies have been successfully produced. In addition, Western blots show that this antibody is specific to the membrane fraction of sweetgum and aspen xylem extract. When the FA5H-2 antibody was added to a reaction mixture containing aspen xylem tissue, enzyme inhibition studies showed that the activity of FA5H in aspen was reduced more than 60%, a further indication that FA5H-2 performs a P450-like function. Fig. 5 (SEQ ID 2) illustrates the FA5H-2 sequence.

Example 5 - Identifying Gymnosperm Promoter Regions

In order to identify gymnosperm promoter regions, sequences from loblolly pine PAL and 4CL1B and 4CL3B lignin genes were used as primers to screen the loblolly pine genomic library, using the GenomeWalker Kit. The loblolly pine PAL primer sequence was obtained from the GenBank, reference number U39792. The loblolly pine 4CL1B primer sequences were also obtained from the gene bank, reference numbers U39404 and U39405.

The loblolly pine genomic library was constructed in Lambda DashII, available from Stratagene, of LaJolla, CA. 3 x 106 phage plaques from the genomic library of loblolly pine were screened using both the above mentioned PAL cDNA and 4CL (PCR clone) fragments as probes. Five 4CL clones were obtained after screening. Lambda DNAs of two 4CL of the five 4CL clones obtained after screening were isolated and digested by EcoRV, PstI, SalI and XbaI for Southern analysis. Southern analysis using 4CL fragments as probes indicated that both clones for the 4CL gene were identical. Results from further mapping showed that none of the original five 4CL clones contained promoter regions. When tested, the PAL clones obtained from the screening also did not contain promoter regions.

In a second attempt to clone the promoter regions associated with the PAL and 4CL a Universal GenomeWalkerTM kit, available from CLONETECH, was used. In the process, total DNA from loblolly pine was digested by several restriction enzymes and ligated into the adaptors (libraries) provied with the kit. Two gene-specific primers for each gene were designed (GSP1 and 2). After two rounds of PCR using these primers and adapter primers of the kit, several fragments were amplified from each library. A 1.6 kb fragment and a 0.6 kb fragment for PAL gene and a 2.3 kb fragment (4CL1B) and a 0.7 kb fragment (4CL3B) for the 4CL gene were cloned, sequenced and found to contain promoter regions for all three genes. See Fig. 6 (SEQ ID 6), 7 (SEQ ID 7) and 8 (SEQ ID 5).

Example 6 - Fusing the ASL DNA Sequence to A Constitutive Promoter Region and Inserting the Expression Cassette Into a Gymnosperm Genome

As a first step, a ASL DNA sequence, FA5H-1, was fused with a constitutive promoter region according to the methods described in the above Section IV to form an FA5H-1 expression cassette. A second ASL DNA sequence, FA5H-2, was then fused with a constitutive promoter in the same manner to form an FA5H-2 expression cassette. The FA5H-1 expression cassette was inserted into the gymnosperm genome by micro-projectile bombaradment. Embryogenic tissue cultures of loblolly pine were initiated from immature zygotic embryos. The tissue was maintained in an undifferentiated state on semi-solid proliferation medium, according to methods described by Newton et al. TAES Technical Publication "Somatic Embryogenesis in Slash Pine", 1995 and Keinonen-Mettala et al. 1996, Scand. J. For. Res. 11: 242-250.

After separation, 5 ml of the liquid cell suspension fraction which passes through the 40 mesh screen was vacuum deposited onto filter paper and placed on semi-solid proliferation medium. The prepared gymnosperm target cells were then grown for 2 days on filter paper discs placed on semi-solid proliferation medium in a petri dish. These target cell were then bombarded with plasmid DNA containing the FA5H-1 expression cassette and an expression cassette containing a selectable marker gene encoding the enzyme which confers resistance to the antibiotic hygromycin B. A 1:1 mixture of of selectable marker expression cassette and plasmid DNA containing the FA5H-1 expression cassette is precipitated with gold (1.5-3.0 microns) as described by Sanford et al. (1992). The DNA-coated microprojectiles were rinsed in absolute ethanol and aliqots of 10 μl (5 μg DNA/3mg gold) were dried onto a macrocarrier, such as those available from BioRad (Hercules, CA).

Prior to bombardment, embryogenic tissue was desiccated under a sterile laminar-flow hood for 5 minutes. The desiccated tissue was transferred to semi-solid proliferation medium. The microprojectiles were accelerated into desiccated target cells using a BioRad PDS-1000/HE particle gun.

Each plate was bombarded once, rotated 180 degrees, and bombarded a second time. Preferred bombardment parameters were 1350 psi rupture disc pressure, 6 mm distance from the rupture disc to macrocarrier (gap distance), 1 cm macrocarrier travel distance, and 10 cm distance from macrocarrier stopping screen to culture plate (microcarrier travel distance). Tissue was then transferred to semi-solid proliferation medium containing hygromycin B for two days after bombardment.

The FA5H-2 expression cassette was inserted into the gymnosperm genome according to the same procedures.

Example 7 - Selecting Transformed Target Cells

After insertion of the FA5H-1 expression cassette and the selectable marker expression cassette into the gymnosperm target cells as described in Example 6, transformed cells were selected by

exposure to an antibiotic that causes mortality of any cells not containing the GSL expression cassette. Forty independent cell lines were established from cultures cobombarded with an expression cassette containing a hygromycin resistance gene construct and the FA5H-1 construct. These cell lines include lines Y2, Y17, Y7 and O4, as discussed in more detail below.

PCR techniques were then used to verify that the FA5H-1 gene had been successfully integrated into the genomes of of the established cell lines by extracting genomic DNA using the Plant DNAeasy kit, available from Quaign. 200 ng DNA from each cell line were used for each PCR reaction. Two FA5H-1 specific primers were designed to perform a PCR reaction with a 600bp PCR product size. The primers were:

LsFa5H-im1-S primer: ATGGCTTTCCTTCTAATACCCATCTC , and

LsFA5H-im1-A primer: GGGTGTAATGGACGAGCAAGGACTTG.

Each PCR reaction (100 μl) consisted of 75 μl H2O, 1 μl MgCl (25 mM), 10 μl PCR buffer 1 μl 10mM dNTPs, and 10 μl DNA. 100 μl oil was layered on the top of each reaction mix. Hot start PCR was done as follows: PCR reaction was incubated at 95 degrees C for 7 minutes and 1 μl each of both LsFA5H-im1-S and LsFa5H-im1-A primers (100 μM stock) and 1 μl of Taq polymerase were added through oil in each reaction. The PCR program used was 95 degrees C for 1.5 minutes, 55 degrees C for 45 sec and 72 degrees C for 2 minutes, repeated for 40 cycles, followed by extension at 72 degrees C for 10 minutes.

The above PCR products were employed to determine if gymnosperm cells contained the angiosperm lignin gene sequences. With reference to Fig. 9, PCR amplification was performed using template DNA from cells which grew vigorously on hygromycin B-containing medium. The PCR products were electrophoresis in an agarose gel containing 9 lanes. Lanes 1-4 contained PCR amplification of products of the Sweetgum FA5H-1 gene from a non-transformed control and transgenic loblolly pine cell lines. Lane 1 contained the non-transformed control PT52. Lane 2 contained transgenic line Y2. Lane 3 contained transgenic line Y17 and Lane 4 contained the plasmid which contains the expression cassette pSSLsFA5H1-im-s. Lanes 2 through 4 all contain an amplified fragment of about 600 bp, indicating that the FA5H-1 gene has been successfully inserted into transgenic cell lines Y2 and Y17.

Lane 5 contained a DNA size marker Phi 174/HaeIII (BRL). The top four bands in this lane indicate molecular sizes of 1353, 1078, 872 and 603 bp.

Lanes 6-9 contained PCR amplification products of hygromycin B gene from non-transformed control and transgenic loblolly pine cell lines. Lane 6 contained the non-transformed control PT52 line, available from _____. Lane 7 contained transgenic line Y7. Lane 8 contained transgenic line O4. Lane 9 contained the plasmid which includes the expression cassette containing the gene encoding the enzyme which confers resistance tot he antibiotic hygromycin B.

Lanes 7-9 all show an amplified fragment of about 1000bp, indicating that the hygromycin gene has been successfully inserted into transgenic lines Y7 and O4.

These PCR results confirmed the presence of FA5H-1 and hygromycin resistance gene in transformed loblolly pine cell cultures. The results obtained from the PCR verification of 4 cell lines, and similar tests with the remaining 36 cell lines, confirm stable integration of the FA5H-1 gene and the hygromycin B gene in 25% of the 40 cell lines.

In addition, loblolly pine embryogenic cells which have been co-bombarded with the FA5H-2 and hygromycin B expression cassettes, are growing vigorously on hygromycin selection medium, indicating that the FA5H-2 expression cassette was successfully integrated into the gymnosperm genome.

Although various embodiments and features of the invention have been described in the foregoing detailed description, those of ordinary skill will recognize the invention is capable of numerous modifications, rearrangements and substitutions without departing from the scope of the invention as set forth in the appended claims. For example, in the case where the lignin DNA sequence is transcribed and translated to produce a functional syringyl lignin gene, those of ordinary skill will recognize that because of codon degeneracy a number of polynucleotide sequences will encode the same gene. These variants are intended to be covered by the DNA sequences disclosed and claimed herein. In addition, the sequences claimed herein include those sequences with encode a gene having substantial functional identity with those claimed. Thus, in the case of syringyl lignin genes, for example, the DNA sequences include variant polynucleotide sequences encoding polypeptides which have substantial identity with the amino acid sequence of syringyl lignin and which show syringyl lignin activity in gymnosperms.

"Sequence Listing"

Applicant:	Chiang, et al
Title of Invention:	PRODUCTION OF SYRINGYL LIGNIN IN GYMNOSPERMS
Number of Sequences:	7
Information for Sequence ID #:	1 (FA5H-1)
Correspondence Address:	Luedeka, Neely & Graham P.O. Box 1871 Knoxville, TN 37901
Addressee:	Mark S. Graham
Street:	P.O. Box 1871
City:	Knoxville
Country:	USA
Zip:	37901
Computer Readable Form:	
Medium Type:	1.44
Operating System:	DOS
Software:	ASCII
Current Application Data:	N/A
Attorney Information:	
Name:	Mark S. Graham
Registration Number:	32,355
Reference/Docket Number:	50617.00
Telecommunication Information:	
Telephone:	(423) 546-4305
Telefax:	(423) 523-4478
Information for Seq ID No:	
Sequence Characteristics:	
Length:	1707
Type:	DNA
Strandedness:	double
Topology:	linear
Molecule Type:	cDNA
Hypothetical:	No
Anti-Sense:	No
Fragment Type:	N/A
Original Source:	
Organism:	*Liquidambar styraciflua* (L.)
Strain:	Wild Type
Individual Isolate:	N/A
Developmental State:	sporophyte
Haplotype:	N/A
Tissue Type:	xylem
Cell Type:	parenchyma
Cell Line:	N/A
Organelle:	N/A

```
CGGCACGAGG AAACCCTAAA ACTCACCTCT CTTACCCTTT CTCTTCA ATG GCT TTC          56
                                                      Met Ala Phe

CTT CTA ATA CCC ATC TCA ATA ATC TTC ATC GTC TTA GCT TAC CAG              101
Leu Leu Ile Pro Ile Ser Ile Ile Phe Ile Val Leu Ala Tyr Gln

CTC TAT CAA CGG CTC AGA TTT AAG CTC CCA CCC GGC CCA CGT CCA              146
Leu Tyr Gln Arg Leu Arg Phe Lys Leu Pro Pro Gly Pro Arg Pro

TGG CCG ATC GTC GGA AAC CTT TAC GAC ATA AAA CCG GTG AGG TTC              191
Trp Pro Ile Val Gly Asn Leu Tyr Asp Ile Lys Pro Val Arg Phe

CGG TGT TTC GCC GAG TGG TCA CAA GCG TAC GGT CCG ATC ATA TCG              236
Arg Cys Phe Ala Glu Trp Ser Gln Ala Tyr Gly Pro Ile Ile Ser

GTG TGG TTC GGT TCA ACG TTG AAT GTG ATC GTA TCG AAT TCG GAA              281
Val Trp Phe Gly Ser Thr Leu Asn Val Ile Val Ser Asn Ser Glu

TTG GCT AAG GAA GTG CTC AAG GAA AAA GAT CAA CAA TTG GCT GAT              326
Leu Ala Lys Glu Val Leu Lys Glu Lys Asp Gln Gln Leu Ala Asp

AGG CAT AGG AGT AGA TCA GCT GCC AAA TTT AGC AGG GAT GGG CAG              371
Arg His Arg Ser Arg Ser Ala Ala Lys Phe Ser Arg Asp Gly Gln

GAC CTT ATA TGG GCT GAT TAT GGA CCT CAC TAT GTG AAG GTT ACA              416
Asp Leu Ile Trp Ala Asp Tyr Gly Pro His Tyr Val Lys Val Thr

AAG GTT TGT ACC CTC GAG CTT TTT ACT CCA AAG CGG CTT GAA GCT              461
Lys Val Cys Thr Leu Glu Leu Phe Thr Pro Lys Arg Leu Glu Ala

CTT AGA CCC ATT AGA GAA GAT GAA GTT ACA GCC ATG GTT GAG TCC              506
Leu Arg Pro Ile Arg Glu Asp Glu Val Thr Ala Met Val Glu Ser

ATT TTT AAT GAC ACT GCG AAT CCT GAA AAT TAT GGG AAG AGT ATG              551
Ile Phe Asn Asp Thr Ala Asn Pro Glu Asn Tyr Gly Lys Ser Met

CTG GTG AAG AAG TAT TTG GGA GCA GTA GCA TTC AAC AAC ATT ACA              596
Leu Val Lys Lys Tyr Leu Gly Ala Val Ala Phe Asn Asn Ile Thr

AGA CTC GCA TTT GGA AAG CGA TTC GTG AAT TCA GAG GGT GTA ATG              641
Arg Leu Ala Phe Gly Lys Arg Phe Val Asn Ser Glu Gly Val Met

GAC GAG CAA GGA CTT GAA TTT AAG GAA ATT GTG GCC AAT GGA CTC              686
Asp Glu Gln Gly Leu Glu Phe Lys Glu Ile Val Ala Asn Gly Leu

AAG CTT GGT GCC TCA CTT GCA ATG GCT GAG CAC ATT CCT TGG CTC              731
Lys Leu Gly Ala Ser Leu Ala Met Ala Glu His Ile Pro Trp Leu

CGT TGG ATG TTC CCA CTT GAG GAA GGG GCC TTT GCC AAG CAT GGG              776
Arg Trp Met Phe Pro Leu Glu Glu Gly Ala Phe Ala Lys His Gly

GCA CGT AGG GAC CGA CTT ACC AGA GCT ATC ATG GAA GAG CAC ACA              821
Ala Arg Arg Asp Arg Leu Thr Arg Ala Ile Met Glu Glu His Thr
```

```
ATA GCC CGT AAA AAG AGT GGT GGA GCC CAA CAA CAT TTC GTG GAT          866
Ile Ala Arg Lys Lys Ser Gly Gly Ala Gln Gln His Phe Val Asp

GCA TTG CTC ACC CTA CAA GAG AAA TAT GAC CTT AGC GAG GAC ACT          911
Ala Leu Leu Thr Leu Gln Glu Lys Tyr Asp Leu Ser Glu Asp Thr

ATT ATT GGG CTC CTT TGG GAT ATG ATC ACT GCA GGC ATG GAC ACA          956
Ile Ile Gly Leu Leu Trp Asp Met Ile Thr Ala Gly Met Asp Thr

ACC GCA ATC TCT GTC GAA TGG GCC ATG GCC GAG TTA ATT AAG AAC         1001
Thr Ala Ile Ser Val Glu Trp Ala Met Ala Glu Leu Ile Lys Asn

CCA AGG GTG CAA CAA AAA GCT CAA GAG GAG CTA GAC AAT GTA CTT         1046
Pro Arg Val Gln Gln Lys Ala Gln Glu Glu Leu Asp Asn Val Leu

GGG TCC GAA CGT GTC CTG ACC GAA TTG GAC TTC TCA AGC CTC CCT         1091
Gly Ser Glu Arg Val Leu Thr Glu Leu Asp Phe Ser Ser Leu Pro

TAT CTA CAA TGT GTA GCC AAG GAG GCA CTA AGG CTG CAC CCT CCA         1136
Tyr Leu Gln Cys Val Ala Lys Glu Ala Leu Arg Leu His Pro Pro

ACA CCA CTA ATG CTC CCT CAT CGC GCC AAT GCC AAC GTC AAA ATT         1181
Thr Pro Leu Met Leu Pro His Arg Ala Asn Ala Asn Val Lys Ile

GGT GGC TAC GAC ATC CCT AAG GGA TCA AAT GTT CAT GTA AAT GTC         1226
Gly Gly Tyr Asp Ile Pro Lys Gly Ser Asn Val His Val Asn Val

TGG GCC GTG GCT CGT GAT CCA GCA GTG TGG CGT GAC CCA CTA GAG         1271
Trp Ala Val Ala Arg Asp Pro Ala Val Trp Arg Asp Pro Leu Glu

TTT CGA CCG GAA CGG TTC TCT GAA GAC GAT GTC GAC ATG AAA GGT         1316
Phe Arg Pro Glu Arg Phe Leu Glu Asp Asp Val Asp Met Lys Gly

CAC GAT TAT AGG CTA CTG CCG TTT GGT GCA GGG AGG CGT GTT TGC         1361
His Asp Tyr Arg Leu Leu Pro Phe Gly Ala Gly Arg Arg Val Cys

CCC GGT GCA CAA CTT GGC ATC AAT TTG GTC ACA TCC ATG ATG GGT         1406
Pro Gly Ala Gln Leu Gly Ile Asn Leu Val Thr Ser Met Met Gly

CAC CTA TTG CAC CAT TTC TAT TGG AGC CCT CCT AAA GGT GTA AAA         1451
His Leu Leu His His Phe Tyr Trp Ser Pro Pro Lys Gly Val Lys

CCA GAG GAG ATT GAC ATG TCA GAG AAT CCA GGA TTG GTC ACC TAC         1496
Pro Glu Glu Ile Asp Met Ser Glu Asn Pro Gly Leu Val Thr Tyr

ATG CGA ACC CCG GTG CAA GCT GTT CCC ACT CCA AGG CTG CCT GCT         1541
Met Arg Thr Pro Val Gln Ala Val Pro Thr Pro Arg Leu Pro Ala

CAC TTG TAC AAA CGT GTA GCT GTG GAT ATG TAATTCTTAG TTTGTTATTA       1591
His Leu Tyr Lys Arg Val Ala Val Asp Met

TTCATGCTCT TAAGGTTTTG GACTTTGAAC TTATGATGAG ATTTGTAAAA TTCCAAGTGA   1651

TCAAATGAAG AAAAGACCAA ATAAAAAGGC TTGACGATTT AAAAAAAAAA AAAAAAA      1708
```

Information for Sequence ID #:	2 (FA5H-2)
Correspondence Address:	Luedeka, Neely & Graham
	P.O. Box 1871
	Knoxville, TN 37901
Addressee:	Mark S. Graham
Street:	P.O. Box 1871
City:	Knoxville
Country:	USA
Zip:	37901
Computer Readable Form:	
Medium Type:	1.44
Operating System:	DOS
Software:	ASCII
Current Application Data:	N/A
Attorney Information:	
Name:	Mark S. Graham
Registration Number:	32,355
Reference/Docket Number:	50617.00
Telecommunication Information:	
Telephone:	(423) 546-4305
Telefax:	(423) 523-4478
Information for Seq ID No:	
Sequence Characteristics:	
Length:	1883
Type:	DNA
Strandedness:	double
Topology:	linear
Molecule Type:	cDNA
Hypothetical:	No
Anti-Sense:	No
Fragment Type:	N/A
Original Source:	
Organism:	*Liquidambar styraciflua* (L.)
Strain:	Wild Type
Individual Isolate:	N/A
Developmental State:	sporophyte
Haplotype:	N/A
Tissue Type:	xylem
Cell Type:	parenchyma
Cell Line:	N/A
Organelle:	N/A

```
TGCAAACCTG CACAAACAAA GAGAGAGAAG AAGAAAAAGG                              40

AAGAGAGGAG AGAGAGAGAG AGAGAGAGAA GC                                      72

CAT GGA TTC TTC TCT TCA TGA AGC CTT GCA ACC ACT ACC CAT GAC GCT          120
Met Asp Ser Ser Leu His Glu Ala Leu Gln Pro Leu Pro Met Thr Leu

GTT CTT CAT TAT ACC TTT GCT ACT CTT ATT GGG CCT AGT ATC TCG GCT          168
Phe Phe Ile Ile Pro Leu Leu Leu Leu Leu Gly Leu Val Ser Arg Leu

TCG CCA GAG ACT ACC ATA CCC ACC AGG CCC AAA AGG CTT ACC GGT GAT          216
Arg Gln Arg Leu Pro Tyr Pro Pro Gly Pro Lys Gly Leu Pro Val Ile

CGG AAA CAT GCT CAT GAT GGA TCA ACT CAC TCA CCG AGG ACT CGC CAA          264
Gly Asn Met Leu Met Met Asp Gln Leu Thr His Arg Gly Leu Ala Lys

ACT CGC CAA ACA ATA CGG CGG TCT ATT CCA CCT CAA GAT GGG ATT CTT          312
Leu Ala Lys Gln Tyr Gly Gly Leu Phe His Leu Lys Met Gly Phe Leu

ACA CAT GGT GGC CGT TTC CAC ACC CGA CAT GGC TCG CCA AGT CCT TCA          360
His Met Val Ala Val Ser Thr Pro Asp Met Ala Arg Gln Val Leu Gln

AGT CCA AGA CAA CAT CTT CTC GAA CCG GCC AGC CAC CAT AGC CAT CAG          408
Val Gln Asp Asn Ile Phe Ser Asn Arg Pro Ala Thr Ile Ala Ile Ser

CTA CCT CAC CTA TGA CCG AGC CGA CAT GGC CTT CGC TCA CTA CGG CCC          456
Tyr Leu Thr Tyr Asp Arg Ala Asp Met Ala Phe Ala His Tyr Gly Pro

GTT TTG GCG TCA GAT GCG TAA ACT CTG CGT CAT GAA ATT ATT TAG CCG          504
Phe Trp Arg Gln Met Arg Lys Leu Cys Val Met Lys Leu Phe Ser Arg

GAA ACG AGC CGA GTC GTG GGA GTC GGT CCG AGA CGA GGT CGA CTC GGC          552
Lys Arg Ala Glu Ser Trp Glu Ser Val Arg Asp Glu Val Asp Ser Ala

AGT ACG AGT GGT CGC GTC CAA TAT TGG GTC GAC GGT GAA TAT CGG CGA          600
Val Arg Val Val Ala Ser Asn Ile Gly Ser Thr Val Asn Ile Gly Glu

GCT GGT TTT TGC TCT GAC GAA GAA TAT TAC TTA CAG GGC GGC TTT TGG          648
Leu Val Phe Ala Leu Thr Lys Asn Ile Thr Tyr Arg Ala Ala Phe Gly

GAC GAT CTC GCA TGA GGA CCA GGA CGA GTT CGT GGC CAT ACT GCA AGA          696
Thr Ile Ser His Glu Asp Gln Asp Glu Phe Val Ala Ile Leu Gln Glu

GTT TTC GCA GCT GTT TGG TGC TTT TAA TAT AGC TGA TTT TAT CCC TTG          744
```

```
Phe Ser Gln Leu Phe Gly Ala Phe Asn Ile Ala Asp Phe Ile Pro Trp

GCT CAA ATG GGT TCC TCA GGG GAT TAA CGT CAG GCT CAA CAA GGC ACG     792
Leu Lys Trp Val Pro Gln Gly Ile Asn Val Arg Leu Asn Lys Ala Arg

AGG GGC GCT TGA TGG GTT TAT TGA CAA GAT CAT CGA CGA TCA TAT ACA     840
Gly Ala Leu Asp Gly Phe Ile Asp Lys Ile Ile Asp Asp His Ile Gln

GAA GGG GAG TAA AAA CTC GGA GGA GGT TGA TAC TGA TAT GGT AGA TGA     888
Lys Gly Ser Lys Asn Ser Glu Glu Val Asp Thr Asp Met Val Asp Asp

TTT ACT TGC TTT TTA CGG TGA GGA AGC CAA AGT AAG CGA ATC TGA CGA     936
Leu Leu Ala Phe Tyr Gly Glu Glu Ala Lys Val Ser Glu Ser Asp Asp

TCT TCA AAA TTC CAT CAA ACT CAC CAA AGA CAA CAT CAA AGC TAT CAT     984
Leu Gln Asn Ser Ile Lys Leu Thr Lys Asp Asn Ile Lys Ala Ile Met

GGA CGT AAT GTT TGG AGG GAC CGA AAC GGT GGC GTC CGC GAT TGA ATG    1032
Asp Val Met Phe Gly Gly Thr Glu Thr Val Ala Ser Ala Ile Glu Trp

GGC CAT GAC GGA GCT GAT GAA AAG CCC AGA AGA TCT AAA GAA GGT CCA    1080
Ala Met Thr Glu Leu Met Lys Ser Pro Glu Asp Leu Lys Lys Val Gln

ACA AGA ACT CGC CGT GGT GGT GGG TCT TGA CCG GCG AGT CGA AGA GAA    1128
Gln Glu Leu Ala Val Val Val Gly Leu Asp Arg Arg Val Glu Glu Lys

AGA CTT CGA GAA GCT CAC CTA CTT GAA ATG CGT ACT GAA GGA AGT CCT    1176
Asp Phe Glu Lys Leu Thr Tyr Leu Lys Cys Val Leu Lys Glu Val Leu

TCG CCT CCA CCC ACC CAT CCC ACT CCT CCT CCA CGA GAC TGC CGA GGA    1224
Arg Leu His Pro Pro Ile Pro Leu Leu Leu His Glu Thr Ala Glu Asp

CGC CGA GGT CGG CGG CTA CTA CAT TCC GGC GAA ATC GCG GGT GAT GAT    1272
Ala Glu Val Gly Gly Tyr Tyr Ile Pro Ala Lys Ser Arg Val Met Ile

CAA CGC GTG CGC CAT CGG CCG GGA CAA GAA CTC GTG GGC CGA CCC AGA    1320
Asn Ala Cys Ala Ile Gly Arg Asp Lys Asn Ser Trp Ala Asp Pro Asp

TAC GTT TAG GCC CTC CAG GTT TCT CAA AGA CGG TGT GCC CGA TTT CAA    1368
Thr Phe Arg Pro Ser Arg Phe Leu Lys Asp Gly Val Pro Asp Phe Lys

AGG GAA CAA CTT CGA GTT CAT CCC ATT CGG GTC AGG TCG TCG GTC TTG    1416
Gly Asn Asn Phe Glu Phe Ile Pro Phe Gly Ser Gly Arg Arg Ser Cys
```

```
CCC CGG TAT GCA ACT CGG ACT CTA CGC GCT AGA GAC GAC TGT GGC TCA   1464
Pro Gly Met Gln Leu Gly Leu Tyr Ala Leu Glu Thr Thr Val Ala His

CCT CCT TCA CTG TTT CAC GTG GGA GTT GCC GGA CGG GAT GAA ACC GAG   1512
Leu Leu His Cys Phe Thr Trp Glu Leu Pro Asp Gly Met Lys Pro Ser

TGA ACT CGA GAT GAA TGA TGT GTT TGG ACT CAC CGC GCC AAG AGC GAT   1560
Glu Leu Glu Met Asn Asp Val Phe Gly Leu Thr Ala Pro Arg Ala Ile

TCG ACT CAC CGC CGT GCC GAG TCC ACG CCT TCT CTG TCC TCT CTA       1605
Arg Leu Thr Ala Val Pro Ser Pro Arg Leu Leu Cys Pro Leu Tyr

TTGATCGAAT GATTGGGGGA GCTTTGTGGA GGGGCTTTTA TGGAGACTCT ATATATAGAT  1665

GGGAAGTGAA ACAACGACAG GTGAATGCTT GGATTTTTGG TATATATTGG GGAGGGAGGG  1725

GAAAAAAAAA ATAATGAAAG GAAAGAAAAG AGAGAATTTG AATTTCTCTT CCTCTGTGGA  1785

TAAAAGCCTC GTTTTTAATT GTTTTTATGT GGAGATATTT GTGTTTGTTT ATTTTTATCT  1845

CTTTTTTTGC AATAACACTC AAAAATAAAA AAAAAAAA                          1883
```

Information for Sequence ID #:	3 (bi-OMT)
Correspondence Address:	Luedeka, Neely & Graham P.O. Box 1871 Knoxville, TN 37901
Addressee:	Mark S. Graham
Street:	P.O. Box 1871
City:	Knoxville
Country:	USA
Zip:	37901
Computer Readable Form:	
Medium Type:	1.44
Operating System:	DOS
Software:	ASCII
Current Application Data:	N/A
Attorney Information:	
Name:	Mark S. Graham
Registration Number:	32,355
Reference/Docket Number:	50617.00
Telecommunication Information:	
Telephone:	(423) 546-4305
Telefax:	(423) 523-4478
Information for Seq ID No:	
Sequence Characteristics:	
Length:	1380
Type:	DNA
Strandedness:	double
Topology:	linear
Molecule Type:	cDNA
Hypothetical:	No
Anti-Sense:	No
Fragment Type:	N/A
Original Source:	
Organism:	*Liquidambar styraciflua* (L.)
Strain:	Wild Type
Individual Isolate:	N/A
Developmental State:	sporophyte
Haplotype:	N/A
Tissue Type:	xylem
Cell Type:	parenchyma
Cell Line:	N/A
Organelle:	N/A

```
CGGCACGAGC CCTACCTCCT TTCTTGGAAA AATTTCCCCA TTCGATCACA ATCCGGGCCT      60

CAAAAA ATG GGA TCA ACA AGC GAA ACG AAG ATG AGC CCG AGT GAA GCA        108
       Met Gly Ser Thr Ser Glu Thr Lys Met Ser Pro Ser Glu Ala

GCA GCA GCA GAA GAA GAA GCA TTC GTA TTC GCT ATG CAA TTA ACC AGT       156
Ala Ala Ala Glu Glu Glu Ala Phe Val Phe Ala Met Gln Leu Thr Ser

GCT TCA GTT CTT CCC ATG GTC CTA AAA TCA GCC ATA GAG CTC GAC GTC       204
Ala Ser Val Leu Pro Met Val Leu Lys Ser Ala Ile Glu Leu Asp Val

TTA GAA ATC ATG GCT AAA GCT GGT CCA GGT GCG CAC ATA TCC ACA TCT       252
Leu Glu Ile Met Ala Lys Ala Gly Pro Gly Ala His Ile Ser Thr Ser

GAC ATA GCC TCT AAG CTG CCC ACA AAG AAT CCA GAT GCA GCC GTC ATG       300
Asp Ile Ala Ser Lys Leu Pro Thr Lys Asn Pro Asp Ala Ala Val Met

CTT GAC CGT ATG CTC CGC CTC TTG GCT AGC TAC TCT GTT CTA ACG TGC       348
Leu Asp Arg Met Leu Arg Leu Leu Ala Ser Tyr Ser Val Leu Thr Cys

TCT CTC CGC ACC CTC CCT GAC GGC AAG ATC GAG AGG CTT TAC GGC CTT       396
Ser Leu Arg Thr Leu Pro Asp Gly Lys Ile Glu Arg Leu Tyr Gly Leu

GCA CCC GTT TGT AAA TTC TTG ACC AGA AAC GAT GAT GGA GTC TCC ATA       444
Ala Pro Val Cys Lys Phe Leu Thr Arg Asn Asp Asp Gly Val Ser Ile

GCC GCT CTG TCT CTC ATG AAT CAA GAC AAG GTC CTC ATG GAG AGC TGG       492
Ala Ala Leu Ser Leu Met Asn Gln Asp Lys Val Leu Met Glu Ser Trp

TAC CAC TTG ACC GAG GCA GTT CTT GAA GGT GGA ATT CCA TTT AAC AAG       540
Tyr His Leu Thr Glu Ala Val Leu Glu Gly Gly Ile Pro Phe Asn Lys

GCC TAT GGA ATG ACA GCA TTT GAG TAC CAT GGC ACC GAT CCC AGA TTC       588
Ala Tyr Gly Met Thr Ala Phe Glu Tyr His Gly Thr Asp Pro Arg Phe

AAC ACA GTT TTC AAC AAT GGA ATG TCC AAT CAT TCG ACC ATT ACC ATG       636
Asn Thr Val Phe Asn Asn Gly Met Ser Asn His Ser Thr Ile Thr Met

AAG AAA ATC CTT GAG ACT TAC AAA GGG TTC GAG GGA CTT GGA TCT GTG       684
Lys Lys Ile Leu Glu Thr Tyr Lys Gly Phe Glu Gly Leu Gly Ser Val

GTT GAT GTT GGT GGT GGC ACT GGT GCC CAC CTT AAC ATG ATT ATC GCT       732
Val Asp Val Gly Gly Gly Thr Gly Ala His Leu Asn Met Ile Ile Ala

AAA TAC CCC ATG ATC AAG GGC ATT AAC TTC GAC TTG CCT CAT GTT ATT       780
Lys Tyr Pro Met Ile Lys Gly Ile Asn Phe Asp Leu Pro His Val Ile

GAG GAG GCT CCC TCC TAT CCT GGT GTG GAG CAT GTT GGT GGA GAT ATG       828
Glu Glu Ala Pro Ser Tyr Pro Gly Val Glu His Val Gly Gly Asp Met

TTT GTT AGT GTT CCA AAA GGA GAT GCC ATT TTC ATG AAG TGG ATA TGT       876
Phe Val Ser Val Pro Lys Gly Asp Ala Ile Phe Met Lys Trp Ile Cys
```

```
CAT GAT TGG AGC GAT GAA CAC TGC TTG AAG TTT TTG AAG AAA TGT TAT        924
His Asp Trp Ser Asp Glu His Cys Leu Lys Phe Leu Lys Lys Cys Tyr

GAA GCA CTT CCA ACC AAT GGG AAG GTG ATC CTT GCT GAA TGC ATC CTC        972
Glu Ala Leu Pro Thr Asn Gly Lys Val Ile Leu Ala Glu Cys Ile Leu

CCC GTG GCG CCA GAC GCA AGC CTC CCC ACT AAG GCA GTG GTC CAT ATT       1020
Pro Val Ala Pro Asp Ala Ser Leu Pro Thr Lys Ala Val Val His Ile

GAT GTC ATC ATG TTG GCT CAT AAC CCA GGT GGG AAA GAG AGA ACT GAG       1068
Asp Val Ile Met Leu Ala His Asn Pro Gly Gly Lys Glu Arg Thr Glu

AAG GAG TTT GAG GCC TTG GCC AAG GGG GCT GGA TTT GAA GGT TTC CGA       1116
Lys Glu Phe Glu Ala Leu Ala Lys Gly Ala Gly Phe Glu Gly Phe Arg

GTA GTA GCC TCG TGC GCT TAC AAT ACA TGG ATC ATC GAA TTT TTG AAG       1164
Val Val Ala Ser Cys Ala Tyr Asn Thr Trp Ile Ile Glu Phe Leu Lys

AAG ATT TGAGTCCTTA CTCGGCTTTG AGTACATAAT ACCAACTCCT TTTGGTTTTC       1220
Lys Ile

GAGATTGTGA TTGTGATTGT GATTGTCTCT CTTTCGCAGT TGGCCTTATG ATATAATGTA    1280

TCGTTAACTC GATCACAGAA GTGCAAAAGA CAGTGAATGT ACACTGCTTT ATAAAATAAA    1340

AATTTTAAGA TTTTGATTCA TGTAAAAAAA AAAAAAAAAA                          1380
```

Information for Sequence ID #:	4 (4CL)
Correspondence Address:	Luedeka, Neely & Graham
	P.O. Box 1871
	Knoxville, TN 37901
Addressee:	Mark S. Graham
Street:	P.O. Box 1871
City:	Knoxville
Country:	USA
Zip:	37901
Computer Readable Form:	
Medium Type:	1.44
Operating System:	DOS
Software:	ASCII
Current Application Data:	N/A
Attorney Information:	
Name:	Mark S. Graham
Registration Number:	32,355
Reference/Docket Number:	50617.00
Telecommunication Information:	
Telephone:	(423) 546-4305
Telefax:	(423) 523-4478
Information for Seq ID No:	
Sequence Characteristics:	
Length:	2026
Type:	DNA
Strandedness:	double
Topology:	linear
Molecule Type:	cDNA
Hypothetical:	No
Anti-Sense:	No
Fragment Type:	N/A
Original Source:	
Organism:	*Liquidambar styraciflua* (L.)
Strain:	Wild Type
Individual Isolate:	N/A
Developmental State:	sporophyte
Haplotype:	N/A
Tissue Type:	xylem
Cell Type:	parenchyma
Cell Line:	N/A
Organelle:	N/A

```
CGGCACGAGC TCATTTTCCA CTTCTGGTTT GATCTCTGCA ATTCTTCCAT CAGTCCCTA        60

ATG GAG ACC CAA ACA AAA CAA GAA GAA ATC ATA TAT CGG TCG AAA              105
Met Glu Thr Gln Thr Lys Gln Glu Glu Ile Ile Tyr Arg Ser Lys

CTC CCC GAT ATC TAC ATC CCC AAA CAC CTC CCT TTA CAT TCG TAT              150
Leu Pro Asp Ile Tyr Ile Pro Lys His Leu Pro Leu His Ser Tyr

TGT TTC GAG AAC ATC TCA CAG TTC GGC TCC CGC CCC TGT CTG ATC              195
Cys Phe Glu Asn Ile Ser Gln Phe Gly Ser Arg Pro Cys Leu Ile

AAT GGC GCA ACG GGC AAG TAT TAC ACA TAT GCT GAG GTT GAG CTC              240
Asn Gly Ala Thr Gly Lys Tyr Tyr Thr Tyr Ala Glu Val Glu Leu

ATT GCG CGC AAG GTC GCA TCC GGC CTC AAC AAA CTC GGC GTT CGA              285
Ile Ala Arg Lys Val Ala Ser Gly Leu Asn Lys Leu Gly Val Arg

CAA GGT GAC ATC ATC ATG CTT TTG CTA CCC AAC TCG CCG GAG TTC              330
Gln Gly Asp Ile Ile Met Leu Leu Leu Pro Asn Ser Pro Glu Phe

GTG TTT TCA ATT CTC GGC GCA TCC TAC CGC GGG GCT GCC GCC ACC              375
Val Phe Ser Ile Leu Gly Ala Ser Tyr Arg Gly Ala Ala Ala Thr

GCC GCA AAC CCG TTT TAT ACC CCT GCC GAG ATC AGG AAG CAA GCC              420
Ala Ala Asn Pro Phe Tyr Thr Pro Ala Glu Ile Arg Lys Gln Ala

AAA ACC TCC AAC GCC AGG CTT ATT ATC ACA CAT GCC TGT TAC TAT              465
Lys Thr Ser Asn Ala Arg Leu Ile Ile Thr His Ala Cys Tyr Tyr

GAG AAA GTG AAG GAC TTG GTG GAA GAG AAC GTT GCC AAG ATC ATA              510
Glu Lys Val Lys Asp Leu Val Glu Glu Asn Val Ala Lys Ile Ile

TGT ATA GAC TCA CCC CCG GAC GGT TGT TTG CAC TTC TCG GAG CTG              555
Cys Ile Asp Ser Pro Pro Asp Gly Cys Leu His Phe Ser Glu Leu

AGT GAG GCG GAC GAG AAC GAC ATG CCC AAT GTA GAG ATT GAC CCC              600
Ser Glu Ala Asp Glu Asn Asp Met Pro Asn Val Glu Ile Asp Pro

GAT GAT GTG GTG GCG CTG CCG TAC TCG TCA GGG ACG ACG GGT TTA              645
Asp Asp Val Val Ala Leu Pro Tyr Ser Ser Gly Thr Thr Gly Leu

CCA AAG GGG GTG ATG CTA ACA CAC AAG GGA CAA GTG ACG AGT GTG              690
Pro Lys Gly Val Met Leu Thr His Lys Gly Gln Val Thr Ser Val

GCG CAA CAG GTG GAC GGA GAG AAT CCG AAC CTG TAT ATA CAT AGC              735
Ala Gln Gln Val Asp Gly Glu Asn Pro Asn Leu Tyr Ile His Ser

GAG GAC GTG GTT CTG TGC GTG TTG CCT CTG TTT CAC ATC TAC TCG              780
Glu Asp Val Val Leu Cys Val Leu Pro Leu Phe His Ile Tyr Ser

ATG AAC GTC ATG TTT TGC GGG TTA CGA GTT GGT GCG GCG ATT CTG              825
Met Asn Val Met Phe Cys Gly Leu Arg Val Gly Ala Ala Ile Leu

ATT ATG CAG AAA TTT GAA ATA TAT GGG TTG TTA GAG CTG GTC AGA              870
Ile Met Gln Lys Phe Glu Ile Tyr Gly Leu Leu Glu Leu Val Arg

AGT ACA GGT GAC CAT CAT GCC TAT CGT ACA CCC ATC GTA TTG GCA              915
```

```
Ser Thr Gly Asp His His Ala Tyr Arg Thr Pro Ile Val Leu Ala

ATC TCC AAG ACT CCG GAT CTT CAC AAC TAT GAT GTG TCC TC ATT          960
Ile Ser Lys Thr Pro Asp Leu His Asn Tyr Asp Val Ser Ser Ile

CGG ACT GTC ATG TCA GGT GCG GCT CCT CTG GGC AAG GAA CT GAA         1005
Arg Thr Val Met Ser Gly Ala Ala Pro Leu Gly Lys Glu Leu Glu

GAT TCT GTC AGA GCT AAG TTT CCC ACC GCC AAA CTT GGT CAGGGA         1050
Asp Ser Val Arg Ala Lys Phe Pro Thr Ala Lys Leu Gly Gln Gly

TAT GGA ATG ACG GAG GCA GGG CCC GTG CTA GCG ATG TGT TTG CA         1095
Tyr Gly Met Thr Glu Ala Gly Pro Val Leu Ala Met Cys Leu Ala

TTT GCC AAG GAA GGG TTT GAA ATA AAA TCG GGG GCA TCT GGA AT         1140
Phe Ala Lys Glu Gly Phe Glu Ile Lys Ser Gly Ala Ser Gly Thr

GTT TTA AGG AAC GCA CAG ATG AAG ATT GTG GAC CCT GAA ACC GG         1185
Val Leu Arg Asn Ala Gln Met Lys Ile Val Asp Pro Glu Thr Gly

GTC ACT CTC CCT CGA AAC CAA CCC GGA GAG ATT TGC ATT AGA GGA        1230
Val Thr Leu Pro Arg Asn Gln Pro Gly Glu Ile Cys Ile Arg Gly

GAC CAA ATC ATG AAA GGT TAT CTT AAT GAT CCT GAG GCG ACG GAG        1275
Asp Gln Ile Met Lys Gly Tyr Leu Asn Asp Pro Glu Ala Thr Glu

AGA ACC ATA GAC AAG GAA GGT TGG TTA CAC ACA GGT GAT GTG GGC        1320
Arg Thr Ile Asp Lys Glu Gly Trp Leu His Thr Gly Asp Val Gly

TAC ATC GAC GAT GAC ACT GAG CTC TTC ATT GTT GAT CGG TTG AAG        1365
Tyr Ile Asp Asp Asp Thr Glu Leu Phe Ile Val Asp Arg Leu Lys

GAA CTG ATC AAA TAC AAA GGG TTT CAG GTG GCA CCC GCT GAG CTT        1410
Glu Leu Ile Lys Tyr Lys Gly Phe Gln Val Ala Pro Ala Glu Leu

GAG GCC ATG CTC CTC AAC CAT CCC AAC ATC TCT GAT GCT GCC GTC        1455
Glu Ala Met Leu Leu Asn His Pro Asn Ile Ser Asp Ala Ala Val

GTC CCA ATG AAA GAC GAT GAA GCT GGA GAG CTC CCT GTG GCG TTT        1500
Val Pro Met Lys Asp Asp Glu Ala Gly Glu Leu Pro Val Ala Phe

GTT GTA AGA TCA GAT GGT TCT CAG ATA TCC GAG GCT GAA ATC AGG        1545
Val Val Arg Ser Asp Gly Ser Gln Ile Ser Glu Ala Glu Ile Arg

CAA TAC ATC GCA AAA CAG GTG GTT TTT TAT AAA AGA ATA CAT CGC        1590
Gln Tyr Ile Ala Lys Gln Val Val Phe Tyr Lys Arg Ile His Arg

GTA TTT TTC GTC GAA GCC ATT CCT AAA GCG CCC TCT GGC AAA ATC        1635
Val Phe Phe Val Glu Ala Ile Pro Lys Ala Pro Ser Gly Lys Ile

TTG CGG AAG GAC CTG AGA GCC AAA TTG GCG TCT GGT CTT CCC AAT        1680
Leu Arg Lys Asp Leu Arg Ala Lys Leu Ala Ser Gly Leu Pro Asn

TAATTCTCAT TCGCTACCCT CCTTTCTCTT ATCATACGCC AACACGAACG AAGAGGCTCA  1740

ATTAAACGCT GCTCATTCGA AGCGGCTCAA TTAAAGCTGC TCATTCATGT CCACCGAGTG  1800
```

GGCAGCCTGT CTTGTTGGGA TGTTCTTTCA TTTGATTCAG CTGTGAGAAG CCAGACCCTC 1860

ATTATTTATT GTGAAATTCA CAAGAATGTC TGTAAATCGA TGTTGTGAGT GATGGGTTTC 1920

AAAACACTTT TGACATTGTT TACGTTGTAT TTCCTGCTGT TGAAAATAAC TACTTTGTAT 1980

GACTTTTATT TGGGAAGATA ACCTTTCAAA AAAAAAAAAA AAAAAA 2026

Information for Sequence ID #:	5 (PAL)
Correspondence Address:	Luedeka, Neely & Graham
	P.O. Box 1871
	Knoxville, TN 37901
Addressee:	Mark S. Graham
Street:	P.O. Box 1871
City:	Knoxville
Country:	USA
Zip:	37901
Computer Readable Form:	
Medium Type:	1.44
Operating System:	DOS
Software:	ASCII
Current Application Data:	N/A
Attorney Information:	
Name:	Mark S. Graham
Registration Number:	32,355
Reference/Docket Number:	50617.00
Telecommunication Information:	
Telephone:	(423) 546-4305
Telefax:	(423) 523-4478
Information for Seq ID No:	
Sequence Characteristics:	
Length:	1544
Type:	DNA
Strandedness:	double
Topology:	linear
Molecule Type:	cDNA
Hypothetical:	No
Anti-Sense:	No
Fragment Type:	N/A
Original Source:	
Organism:	Pinus taeda (L.)
Strain:	Wild Type
Individual Isolate:	N/A
Developmental State:	sporophyte
Haplotype:	N/A
Tissue Type:	xylem
Cell Type:	parenchyma
Cell Line:	N/A
Organelle:	N/A

AAAGATAATA TATGTGTATG CCTACTACTA CACATTGTTT TGAAGTGTGT AAACATAGTG 60

CAACACTAGG AGGACTCACA ATGAGCACTT GTTGACATGA AACTAGCTAA ATGCCCAACA 120

ATATTAGTGA AAGCTAGTTA AACTAACCCC TTTGACTTTC AAGATGATAT ATTTATATCC 180

CTACTACGTC TTCCTCTTTT TGTCTTTCTC TTGTGATTAA ACCTTCCTTG AAACAATTCT 240

CAAATGTAAA ATTAAACCTT GAAACTTGTA GAGACCAAAC TTCCCTAGGA GAAACCACAT 300

TTATGACAAC ATATATACAC CAACCCATTG CATACTATAA TATTGGAATT ACCTGCAGCG 360

AACGAAAGAA ACGCTGTCTC ACCAACTCGT GCACTACATC CCGAAACTTA ACCTTCCCCT 420

GATACAGATT GAAGAGCCGA AAAAAGCGTG CATCCAAATT TCTGGTATGG TGAGGAGCCG 480

AAAAACGCGT GCGCCTAATT TTTTTGAGAT GGGCCGGAAA ATAATGCGTG CATCTAAATT 540

TTCACGTGTC GCGTATTGGC GAGGTTGCGC TGAATGTGAT CCTGTGCGTG AGCCACATTC 600

ATTCCATTGG TTGACCCGCC GGTACCGCGA GGACCGTGGG GTCTCACAGA TACGCGGATG 660

GTGGATCAGC ACTGAGAAGA TTAGATGATG ACCAGGCGGG CATTTGAAGT AAAAACTTGG 720

GGGTGGTTGG CAAGTACGCG ACAAAGAGGG GTAGTGCGCA AGGAAGCGAG TTGGATGCAA 780

ATAATATTAC AAAGTGGGTT GGTGGGCATG AGCATCAACC AGAATGATGT TGTTGCTGGT 840

TCCGTGCAAA TTCTGACCAG TAGTTTGAAC AATACTACCC AACTTGTTTT TGGTAAAACA 900

TGAAGTGGGT AAGGAGAATT GAACTTACGT CTCATGGTAA AGGGCAAGGG CAAATGACTT 960

AACACATACC TTTAACTAAT AAAAATACCC CTAACAAATA CGAAAACGAA TGAGTTATCA 1020

CAGACCTTCA ACTAATAAGA TAGCCATCAG ACCCACATCT CCTGACTGAC CAAAAACAAA 1080

TGACTTCAAC CAACTAAGAT ACCCATCAAA GCTAACCCAC AACCCAATTC CTCACTTCCC 1140

CTTACCAGAC CAACCAAGCA GACCTACGCC ATTAACTACT TTAGGACGTG GGAATTGGGG 1200

GTGCCACCGT TGAAGAATGG CACTCAGGGT TGGTAATCCC TCCACGTGTA TGTAGCAGTC 1260

GTTTGGTGGA GACGGCGTGT TTGAATGTCC ACCTTCCAGT TTGGAGAACA AGGAAATTGG 1320

GCTTATATTA GGCCTGGATC TCTTGTTTCA GAGCAGGAGT AGTTCAGGAC AGGAACTAGC 1380

ATTCAAGAAT TCAATTGCCC TGCCCTGCTC TGCTCTGCTT TGCTCAACTT ATTGATCCCT 1440

GCTCTGGTTT GTTCAATTTC TTGACCCCTG CTGGGTTCTG CTCTGGTTTG CACACTTTCT 1500

CGATTATATA AGTCATTTTG GATCCTTGCA AGGAAGAGAA TATG 1544

Information for Sequence ID #:	6 (4CL1B)
Correspondence Address:	Luedeka, Neely & Graham
	P.O. Box 1871
	Knoxville, TN 37901
Addressee:	Mark S. Graham
Street:	P.O. Box 1871
City:	Knoxville
Country:	USA
Zip:	37901
Computer Readable Form:	
Medium Type:	1.44
Operating System:	DOS
Software:	ASCII
Current Application Data:	N/A
Attorney Information:	
Name:	Mark S. Graham
Registration Number:	32,355
Reference/Docket Number:	50617.00
Telecommunication Information:	
Telephone:	(423) 546-4305
Telefax:	(423) 523-4478
Information for Seq ID No:	
Sequence Characteristics:	
Length:	659
Type:	DNA
Strandedness:	double
Topology:	linear
Molecule Type:	cDNA
Hypothetical:	No
Anti-Sense:	No
Fragment Type:	N/A
Original Source:	
Organism:	Pinus taeda (L.)
Strain:	Wild Type
Individual Isolate:	N/A
Developmental State:	sporophyte
Haplotype:	N/A
Tissue Type:	xylem
Cell Type:	parenchyma
Cell Line:	N/A
Organelle:	N/A

```
AAACACCAAT TTAATGGGAT TTCAGATTTG TATCCCATGC TATTGGCTAA GGCATTTTTC      60

TTATTGTAAT CTAACCAATT CTAATTTCCA CCCTGGTGTG AACTGACTGA CAAATGCGGT     120

CCGAAAACAG CGAATGAAAT GTCTGGGTGA TCGGTCAAAC AAGCGGTGGG CGAGAGAGCG     180

CGGGTGTTGG CCTAGCCGGG ATGGGGGTAG GTAGACGGCG TATTACCGGC GAGTTGTCCG     240

AATGGAGTTT TCGGGGTAGG TAGTAACGTA GACGTCAATG GAAAAAGTCA TAATCTCCGT     300

CAAAAATCCA ACCGCTCCTT CACATCGCAG AGTTGGTGGC CACGGGACCC TCCACCCACT     360

CACTCAATCG ATCGCCTGCC GTGGTTGCCC ATTATTCAAC CATACGCCAC TTGACTCTTC     420

ACCAACAATT CCAGGCCGGC TTTCTATACA ATGTACTGCA CAGGAAAATC CAATATAAAA     480

AGCCGGCCTC TGCTTCCTTC TCAGTAGCCC CCAGCTCATT CAATTCTTCC CACTGCAGGC     540

TACATTTGTC AGACACGTTT TCCGCCATTT TTCGCCTGTT TCTGCGGAGA ATTTGATCAG     600

GTTCGGATTG GGATTGAATC AATTGAAAGG TTTTTATTTT CAGTATTTCG ATCGCCATG      659
```

Information for Sequence ID #:	7 (4CL3B)
Correspondence Address:	Luedeka, Neely & Graham
	P.O. Box 1871
	Knoxville, TN 37901
Addressee:	Mark S. Graham
Street:	P.O. Box 1871
City:	Knoxville
Country:	USA
Zip:	37901
Computer Readable Form:	
Medium Type:	1.44
Operating System:	DOS
Software:	ASCII
Current Application Data:	N/A
Attorney Information:	
Name:	Mark S. Graham
Registration Number:	32,355
Reference/Docket Number:	50617.00
Telecommunication Information:	
Telephone:	(423) 546-4305
Telefax:	(423) 523-4478
Information for Seq ID No:	
Sequence Characteristics:	
Length:	2251
Type:	DNA
Strandedness:	double
Topology:	linear
Molecule Type:	cDNA
Hypothetical:	No
Anti-Sense:	No
Fragment Type:	N/A
Original Source:	
Organism:	Pinus taeda (L.)
Strain:	Wild Type
Individual Isolate:	N/A
Developmental State:	sporophyte
Haplotype:	N/A
Tissue Type:	xylem
Cell Type:	parenchyma
Cell Line:	N/A
Organelle:	N/A

GGCCGGGTGG TGACATTTAT TCATAAATTC ATCTCAAAAC AAGAAGGATT TACAAAAATA 60

AAAGAAAACA AAATTTTCAT CTTTAACATA ATTATAATTG TGTTCACAAA ATTCAAACTT 120

AAACCCTTAA TATAAAGAAT TTCTTTCAAC AATACACTTT AATCACAACT TCTTCAATCA 180

CAACCTCCTC CAACAAAATT AAAATAGATT AATAAATAAA TAAACTTAAC TATTTAAAAA 240

AAAATATTAT ACAAAATTTA TTAAAACTTC AAAATAAACA AACTTTTTAT ACAAAATTCA 300

TCAAAACTTT AAAATAAAGC TAAACACTGA AAATGTGAGT ACATTTAAAA GGACGCTGAT 360

CACAAAAATT TTGAAAACAT AAACAAACTT GAAACTCTAC CTTTTAAGAA TGAGTTTGTC 420

GTCTCATTAA CTCATTAGTT TTATAGTTCG AATCCAATTA ACGTATCTTT TATTTTATGG 480

AATAAGGGTG TTTTAATAAG TGATTTTGGG ATTTTTTTAG TAATTTATTT GTGATATGTT 540

ATGGAGTTTT TAAAAATATA TATATATATA TATATTTTTG GGTTGAGTTT ACTTAAAATT 600

TGGAAAAGGT TGGTAAGAAC TATAAATTGA GTTGTGAATG AGTGTTTTAT GGATTTTTTA 660

AGATGTTAAA TTTATATATG TAATTAAAAT TTTATTTTGA ATAACAAAAA TTATAATTGG 720

ATAAAAAATT GTTTTGTTAA ATTTAGAGTA AAAATTTCAA AATCTAAAAT AATTAAACAC 780

TATTATTTTT AAAAAATTTG TTGGTAAATT TTATCTTATA TTTAAGTTAA AATTTAGAAA 840

AAATTAATTT TAAATTAATA AACTTTTGAA GTCAAATATT CCAAATATTT TCCAAAATAT 900

TAAATCTATT TTGCATTCAA AATACAATTT AAATAATAAA ACTTCATGGA ATAGATTAAC 960

CAATTTGTAT AAAAACCAAA AATCTCAAAT AAAATTTAAA TTACAAAACA TTATCAACAT 1020

TATGATTTCA AGAAAGACAA TAACCAGTTT CCAATAAAAT AAAAAACCTC ATGGCCCGTA 1080

ATTAAGATCT CATTAATTAA TTCTTATTTT TTAATTTTTT TACATAGAAA ATATCTTTAT 1140

ATTGTATCCA AGAAATATAG AATGTTCTCG TCCAGGGACT ATTAATCTCC AAACAAGTTT 1200

CAAAATCATT ACATTAAAGC TCATCATGTC ATTTGTGGAT TGGAAATTAT ATTGTATAAG 1260

AGAAATATAG AATGTTCTCG TCTAGGGACT ATTAATTTCC AAACAAATTT CAAAATCATT 1320

ACATTAAAGC TCATCATGTC ATTTGTGGAT TGGAAATTAG ACAAAAAAAA TCCCAAATAT 1380

TTCTCTCAAT CTCCCAAAAT ATAGTTCGAA CTCCATATTT TTGGAAATTG AGAATTTTTT 1440

TACCCAATAA TATATTTTTT TATACATTTT AGAGATTTTC CAGACATATT TGCTCTGGGA 1500

TTTATTGGAA TGAAGGTTGA GTTATAAACT TTCAGTAATC CAAGTATCTT CGGTTTTTGA 1560

AGATACTAAA TCCATTATAT AATAAAAACA CATTTTAAAC ACCAATTTAA TGGGATTTCA 1620

GATTTGTATC CCATGCTATT GGCTAAGGCA TTTTTCTTAT TGTAATCTAA CCAATTCTAA 1680

TTTCCACCCT GGTGTGAACT GACTGACAAA TGCGGTCCGA AAACAGCGAA TGAAATGTCT 1740

GGGTGATCGG TCAAACAAGC GGTGGGCGAG AGAGCGCGGG TGTTGGCCTA GCCGGGATGG 1800

GGGTAGGTAG ACGGCGTATT ACCGGCGAGT TGTCCGAATG GAGTTTTCGG GGTAGGTAGT 1860

AACGTAGACG TCAATGGAAA AAGTCATAAT CTCCGTCAAA AATCCAACCG CTCCTTCACA 1920

TCGCAGAGTT GGTGGCCACG GGACCCTCCA CCCACTCACT CGATCGCCTG CCGTGGTTGC 1980

CCATTATTCA ACCATACGCC ACTTGACTCT TCACCAACAA TTCCAGGCCG GCTTTCTATA 2040

CAATGTACTG CACAGGAAAA TCCAATATAA AAAGCCGGCC TCTGCTTCCT TCTCAGTAGC 2100

CCCCAGCTCA TTCAATTCTT CCCACTGCAG GCTACATTTG TCAGACACGT TTTCCGCCAT 2160

TTTTCGCCTG TTTCTGCGGA GAATTTGATC AGGTTCGGAT TGGGATTGAA TCAATTGAAA 2220

GGTTTTTATT TTCAGTATTT CGATCGCCAT G 2251